In Proc. of IEEE ICRA Workshop on Semantic Perception, Mapping, and Exploration, Saint Paul, MN, USA, May, 2012

Place Recognition using Surface Entropy Features

Torsten Fiolka[1], Jörg Stückler[2], Dominik A. Klein[3],
Dirk Schulz[1], and Sven Behnke[2]

***Abstract*— In this paper, we present an interest point detector and descriptor for 3D point clouds and depth images, coined SURE, and use it for recognizing semantically distinct places in indoor environments. We propose an interest operator that selects distinctive points on surfaces by measuring the variation in surface orientation based on surface normals in the local vicinity of a point. Furthermore, we design a view-pose-invariant descriptor that captures local surface properties and incorporates colored texture information. In experiments, we compare our approach to a state-of-the-art feature detector in depth images (NARF). Our descriptor achieves superior results for matching interest points between images and also requires lower computation time. Finally, we evaluate the use of SURE features for recognizing places.**

***Index Terms*— surface interest points, local shape-texture descriptor, place recognition**

I. Introduction

Interest points paired with a descriptor of local image context provide a compact representation of image content. Applications such as place or object recognition require that a detector repeatably finds interest points across images taken from various view poses and under differing lighting conditions. Descriptors, on the other hand, are designed to distinguish well between different shapes and textures. However, one must admit that descriptor distinctiveness depends clearly on the variety of shapes and textures that appear at the selected interest points. Thus, a detector will be preferable, if it selects interest points in various structures and highly expressive regions.

In this paper, we propose an approach for extracting shape features at surface points through a measure of surface entropy (SURE). We demonstrate the use of SURE features for place recognition. Our features consist of a novel pair of interest point detector and local context description. Our approach can be applied to depth images as well as unorganized 3D point clouds. An entropy-based interest measure selects points on surfaces that exhibit strong local variation in surface orientation. We complete our approach by the design of a descriptor that captures local surface curvature properties. We also incorporate color and texture cues into the descriptor in case RGB information is available for the points.

[1] Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Wachtberg, Germany torsten.fiolka at fkie.fraunhofer.de, dirk.schulz at fkie.fraunhofer.de

[2] Autonomous Intelligent Systems, Computer Science Institute VI, University of Bonn, Germany stueckler at ais.uni-bonn.de, behnke at cs.uni-bonn.de

[3] Intelligent Vision Systems, Computer Science Institute III, University of Bonn, Germany kleind at iai.uni-bonn.de

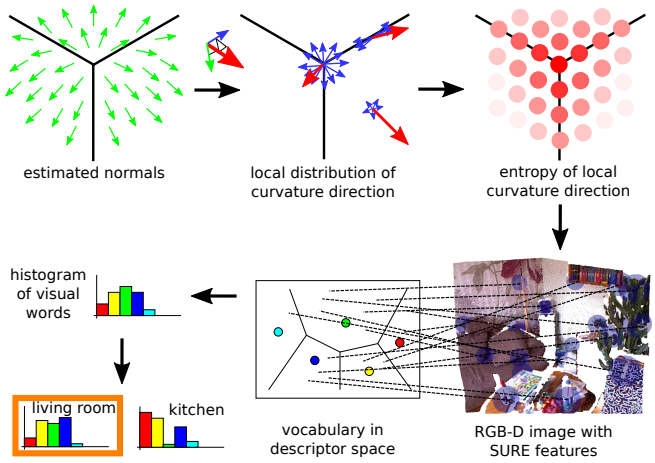


Fig. 1: We detect SURE features in depth images at locations with locally prominent surface curvature. Our interest operator measures the entropy of the distribution of curvature directions at a point in a local neighborhood (top row). The curvature direction (blue arrow) is obtained by the cross product between the estimated surface normal at the point of interest (red arrows) and at neighboring points (green arrows). We propose a descriptor that captures local shape and colored texture at interest points. We recognize places using a Bag-of-Words approach using SURE features (bottom row).

In experiments, we measure repeatability of our interest points under view pose changes for several scenes and objects and compare our approach with a state-of-the-art detector and descriptor to demonstrate advantages of our approach. We show that SURE is capable of correctly recognizing the semantic label of scenes with a Bag-of-Words approach. The top row in Fig. 1 gives a short idea how the interest point detection works while the bottom row outlines the place recognition application with SURE features.

II. Related Work

A. Interest Point Detection

Feature detection and description has been a very active area of research since decades. Nowadays, interest point detection algorithms are designed to be invariant against moderate scale and viewpoint changes [1]. Examples are the Harris-Affine [2] detector that recognizes corner structures based on the second moment matrix, the MSER [3] detector that identifies groups of pixels that are best separable from their surrounding, and the well known SIFT [4] or optimized

SURF [5] detectors that are based on intensity blobs found by a difference of Gaussians filter. Most related to our method, also the entropy measure based on image intensities has been investigated for interest point detection [6], [7], [8]. It has been successfully applied to object recognition [9] due to the high informativeness of maximum entropy regions.

However, those methods purely based on intensity image data suffer problems emerging from projective reduction to 2D space [10]. Recently, various methods have been developed to extract interest points from dense, full-view point clouds.

Novatnack et al. [11] extract multi-scale geometric interest points from dense point clouds with an associated triangular connectivity mesh. Our approach does not require connectivity information given by a mesh. Unnikrishnan et al. [12] derive an interest operator and a scale selection scheme for unorganized point clouds. They extract geodesic distances between points using disjoint minimum spanning trees in a time-consuming pre-processing stage. In [13], this approach has been applied to depth images and an interest detector for corners with scale selection has been proposed. Steder et al. [14] extract interest points from depth images without scale selection, based on a measure of principal curvature which they extent to depth discontinuities. However, our approach is not restricted to depth images and can be readily employed for full-view point clouds.

B. Local Descriptors

The SIFT-descriptor [4] has been successfully used in computer vision applications, as have several improvements to SIFT. SURF [5] sums Haar wavelet responses as a representation of the local gradient pattern. Recently, Calonder et al. [15] and Rublee et al. [16] demonstrated that binarized pixel comparisons yield a robust and highly efficient descriptor.

Several point descriptors have been proposed for 3D point clouds and depth images. Prominent examples are spin-images [17], shape context [18], [19], and (C-)SHOT [20]. Steder et al. [14] proposed the NARF descriptor for depth images. They determine a dominant orientation from depth gradients in a local image patch and extract radial depth gradient histograms. We directly compare our approach to NARF and demonstrate that SURE finds more distinct features. Rusu et al. [21] quantify local surface curvature in rotation-invariant Fast Point Feature Histograms (FPFH). They demonstrate that the histograms can well distinguish between shapes such as corners, spheres, and edges. We also base our descriptor on surfel-pair relations to describe shape. We complement this descriptor with colorful texture cues.

C. Place Recognition

Bag-of-Words approaches are frequently used for place recognition purposes [22], [23]. The image content is first compressed into a set of features that are then quantized using a vocabulary of visual words. The words represent clusters in descriptor space that are learned from training images. The histogram of word occurrences in an image is then compared to the training images in order to retrieve the place category. We also follow a Bag-of-Words approach and demonstrate the use of SURE features for place recognition.

III. Entropy-based Interest Points in 3D Point Clouds

A. Interest Points of Local Surface Entropy

Our detector is based on statistics about the distribution of local surface orientations. We are interested in regions with maximal diversely oriented surfaces, since they show promise to be stably located at transitions of multiple surfaces or capture entire (sub-)structures that stick out of the surroundings. To identify such regions, we measure the entropy

$$H(X_{\mathcal{E}}) = -\sum_{x \in X_{\mathcal{E}}} p(x) \log p(x), \tag{1}$$

where $X_{\mathcal{E}}$ is a random variable characterizing the distribution of surface orientations occurring within a region of interest $\mathcal{E} \subseteq \mathbb{R}^3$. We extract interest points where this entropy measure achieves local maxima, i.e. where $X_{\mathcal{E}}$ is most balanced.

B. Estimation of Surface Normals and Curvature Directions

As we make use of surface normals to estimate surface orientations, we shortly introduce our approach to estimate them. Depth sensors usually measure surfaces by a set of discrete sample points $Q = \{\vec{q}_1, \ldots, \vec{q}_n\}, \vec{q}_k \in \mathbb{R}^3$. We approximate the surface normal at a sample point $n(\vec{q}_k)$ looking at the subset of neighboring points $\mathcal{N}_k = \{\vec{q}_l \in Q | \, \|\vec{q}_k - \vec{q}_l\|_1 < r\}$ within a given support range r. Then, $\hat{n}_r(\vec{q}_k)$ equals the eigenvector corresponding to the smallest eigenvalue of the sample covariance matrix $\mathrm{cov}(\mathcal{N}_k)$.

When directly measuring surface variation from the orientation of surface normals, measurement noise may cause spurious detections at ridges and edges. Heuristic post-processing would then be required to filter out such false detections. Instead, we propose to measure the variation in surface curvature by the cross-product of pairs of surface normals. This second-order statistics exhibit strong orientation peaks and therefore low entropy for ridge- and edge-like structures. In contrast, at structures with diverse surface normal orientations such as corners, curvature directions will also point in diverse directions and entropy will be high. On planar surfaces, curvature direction is strongly affected by noise and, hence, the entropy is meaningless. Nevertheless, the planarity of the surface is indicated by the similarity of the normal orientations. We thus add a new class for parallel normal pairs and re-weight the curvature direction according to the scalar product of the normals.

For every sample point $\vec{q}_k$ we collect all estimated normals $\hat{N}_{\vec{q}_k}$ in the neighborhood $\mathcal{N}$ and calculate the main normal $\hat{n}(\mathcal{E})$, which incorporates all points in $\mathcal{E}$. We calculate curvature directions by the normalized cross products $c_k = (\hat{n}(\mathcal{E}) \times \hat{n}_r(\vec{q}_k)) \,/\, \|\hat{n}(\mathcal{E}) \times \hat{n}_r(\vec{q}_k)\|_2$ between the main normal and the neighboring normals. In order to handle planar surfaces, we assign a weight

$$w_{\times} = (1 - \langle \hat{n}(\mathcal{E}), \hat{n}_r(\vec{q}_k) \rangle) \tag{2}$$

which indicates the likelihood of the normal pair belonging to a planar surface or not.

C. Entropy Calculation

We discretize the domain of weighted curvature directions into a fixed number of orientation bins and a center bin for parallel surfel-pairs. For the discretization of the orientation distribution in a given region of the surface we estimate an orientation histogram. We use the approach by Shah [24] for subdividing a spherical surface into approximately equally sized patches. Every patch is specified by its central azimuth and inclination angle. The number of azimuth angles is determined by the inclination angle so that it is proportional to the circumference of the section of the sphere. We transform the angles from spherical into Cartesian coordinates and obtain a set of normalized vectors $\vec{v}_{i,j}$ pointing to the centers of histogram bins.

A normalized vector $\vec{v}$ contributes to the orientation histogram bin $h_{i,j}$ with a weight inversely proportional to its distance from the center of a histogram bin, i. e.,

$$w_{i,j} = \begin{cases} 0 & \text{, if } \langle \vec{v}, \vec{v}_{i,j} \rangle < \cos\alpha \\ \frac{\langle \vec{v}, \vec{v}_{i,j} \rangle - \cos\alpha}{1-\cos\alpha} & \text{, otherwise.} \end{cases} \quad (3)$$

The maximal angular range of influence is bounded by α. In addition, we weight each entry in the orientation histogram with the unplanarity weight $w_\times$.

The center bin receives the weight $1-w_\times$ for each normal-pair. Finally, we normalize the complete histogram before calculating the entropy according to Equation 1.

D. Efficient Implementation using Octrees

For efficient data access and well-ordered computation, we set up an octree structure containing the 3D point data from the depth image or point cloud. In order to measure local surface entropy, our octree enables uniform sampling in 3D space. Furthermore, we exploit the multi-resolution architecture of the octree for fast volume queries of point statistics.

The multi-scale structure of the octree allows for efficient bottom-up integration of data, facilitating the calculation of histograms, as well as search queries for local maxima in arbitrary volumes. In each node, we store histogram, integral and maximum statistics for different attributes of all points that are located within the volume of the node. These values can be computed efficiently by passing the attributes of points on a path from leaf nodes to the root of the tree. This direction, every parent node accumulates and merges data received from its child nodes. An easily understood example for data statistics is the average position of points within a certain volume $\mathcal{V}$. By integrating over the homogeneous coordinates of points $\vec{s} = (x, y, z, w)^T = \sum_{\vec{q}_i \in \mathcal{V}} (x_i, y_i, z_i, 1)^T$, one retains the mean via normalization $\vec{q} = \frac{1}{w}\vec{s}$.

When querying for statistics inside an arbitrary 3D volume, we recursively descend the tree: if a node is fully inside the queried volume, its statistics are integrated into the response; if it is completely outside, this branch is discontinued; otherwise its child nodes are examined the same way. This is valid since each node already integrates the data of all leaves below in its own statistics.

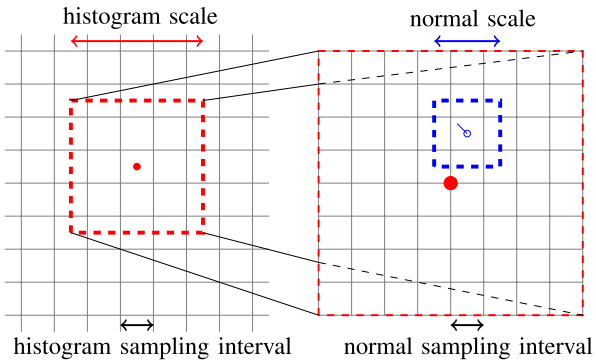


Fig. 2: We calculate normals and entropy histograms at equidistant sample points in fixed scale radii.

E. Interest Point Detection

The surface entropy function depends on two scale parameters: one is the radius r of vicinity $\mathcal{N}$ for the estimation of a surface normal orientation (called normal scale); the other is the extend of a region of interest $\mathcal{E}$, where the distribution of normals and thus the local surface entropy is gathered (called histogram scale) as seen in Fig. 2. These volumes are chosen to be cubic and appropriate to fit the intrinsic octree resolutions. The maximal depth ($\widehat{=}$resolution) of the octree is usually determined by the normal sampling interval at the finest scale, that is specified to be a common multiple of the other dimensions. This way, range queries are processed most efficiently. Usually, sampling interval sizes of surface normals as well as normal orientation histograms are set to be at least half of the diameter of their respective local support volume.

All these parameters have to be chosen carefully. The histogram scale $\mathcal{E}$ corresponds directly to the size of the interest points, at which local structures become salient. Its sampling interval is a trade-off between preciseness and speed. According to the Nyquist-Shannon sampling theorem, a minimal sampling frequency of twice the region size is needed to reconstruct the surface entropy function, i.e. not to miss the occurrence of a local maximum. We choose the normal scale r to a constant fraction of the histogram scale. Accordingly, the sampling interval for normals must also obey the sampling theorem. Reproducing the effect of a lowpass filter for removal of artifacts, we consider an entropy sample to be an interest point candidate, if it exceeds all its spatial neighbors within a dominance region. In addition, the candidate is only kept if it exceeds a global entropy threshold H_{min}. The latter is checked, because we assume regions of interest containing many equally aligned normals to be unstable and/or less accurate.

F. Improved Localization

Since the detector so far only considers a fixed discretization with an arbitrary global shift, the true maximum location

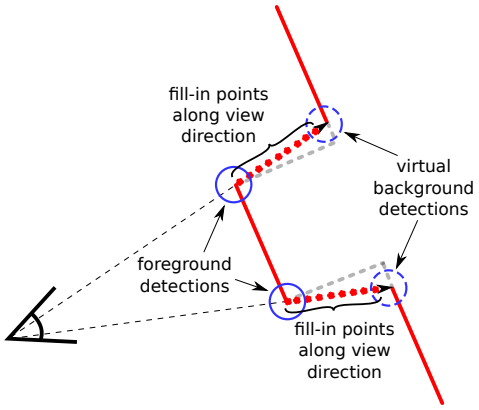


Fig. 3: Occlusion handling. In depth images, structure may be occluded (dashed gray). At depth discontinuities, we therefore add artificial measurements (red dots) from foreground towards the background. We reject false interest point detections at virtual structure in the background.

in entropy has not yet been recovered. In order to improve localization, we apply mean-shift starting from a candidate's location: We integrate surrounding surface entropy samples via a Gaussian window in order to estimate the gradient of the surface entropy density. Then, the position of the candidate is shifted along this gradient direction. We iterate this procedure up to three times.

G. Occlusion Handling in Depth Images

In depth images, one cannot always measure all joining surfaces explicitly due to occlusions, resulting in a reduced entropy. To compensate this we detect jump edges in the depth image. Since we know that there must exist another hidden surface behind each foreground edge, we approximate it by adding artificial measurements in viewing direction up to a distance that meets the biggest used histogram scale. A scheme is show in Fig. 3 We exclude interest points at partial occlusions in the background, since one cannot make any assumptions on the occluded surfaces.

IV. Local Shape-Texture Descriptor

Since our detector finds interest points at locations where the surface exhibits strong local curvature variation, we design a shape descriptor to capture this distribution. When RGB information is available, we also describe the local texture at an interest point. We aim at a rotation-invariant description of the interest points in order to match features despite of view pose changes. For each individual cue, we select a reasonable distance metric and combine them in a distance measure for the complete feature.

A. Shape

Surfel-pair relations have been demonstrated to be a powerful feature for describing local surface curvature [25], [21] (see Fig. 4). In order to describe curvature in the local vicinity of an interest point, we build histograms of surfel-pair relations from neighboring surfels (see Fig. 5). Each surfel is related to the surfel at the interest point being the

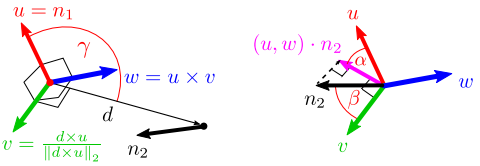


Fig. 4: Surfel-pair relations describe rotation-invariant relative orientations and distances between two surfels.

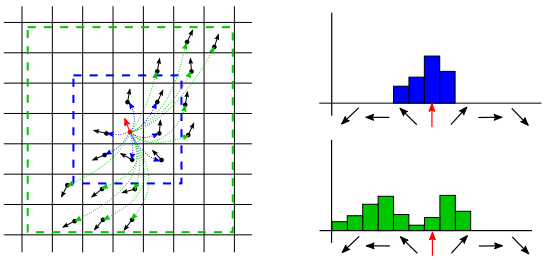


Fig. 5: Shape descriptor in a simplified 2D example. We build histograms of surfel-pair relations from the surfels in a local neighborhood at an interest point. We relate surfels to the central surfel at the interest point. Histograms of inner and outer volumes capture distance-dependent curvature changes.

reference surfel (p_1, n_1). We discretize the angular features into 11 bins each, while we use 2 distance bins to describe curvature in inner and outer volumes. We choose the support size of the descriptor in proportion to the histogram scale.

B. Color

A good color descriptor should allow interest points to be matched despite illumination changes. We choose the HSL color space and build histograms over hue and saturation in the local context of an interest point (see Fig. 6). Our histograms contain 24 bins for hue and one bin for unsaturated, i.e., "gray", colors. Each entry to a hue bin is weighted with the saturation s of the color. The gray bin receives a value of $1-s$. In this way, our histograms also capture information on colorless regions.

Similar to the shape descriptor, we divide the descriptor into 2 histograms over inner and outer volumes at the interest point. In this way, we measure the spatial distribution of color but still retain rotation-invariance.

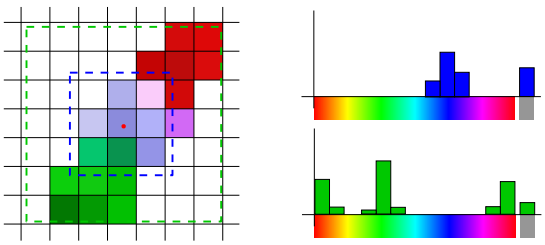


Fig. 6: Color descriptor. We extract hue and saturation histograms in an inner and outer local volume at an interest point.

C. Luminance

Since the color descriptor cannot distinguish between black and white, we propose to quantify luminance contrasts of neighboring points towards the interest point. By this, our luminance descriptor is still invariant to ambient illumination. We use 10 bins for the relative luminance and, again, extract 2 histograms in inner and outer volumes.

D. Measuring Descriptor Distance

The character of the individual components of our descriptor suggests different kinds of distance metrics. For the shape descriptor, we use the Euclidean distance as proposed for FPFH features in [21]. Since the HSL color space is only approximately illumination invariant, the domains of our color histograms may shift and may slightly be misaligned between frames. Hence, the Euclidean distance is not suitable. Instead, we apply an efficient variant of the Earth Mover's Distance (EMD, [26]) from [27] which has been shown to be a robust distance measure on color histograms.

V. Place Recognition

We use SURE features for place recognition in indoor environments. Our approach has two stages: For training, we extract SURE features from each frame $f \in F_{train}$ in a training set of images. We apply the k-means algorithm on the descriptors of the SURE features to create a Bag-of-Words B (BoW) with $k = 400$ visual words. For each frame, we then build similarity histograms $f \in F_{train}$ using the BoW. In the recall stage, we again create similarity histograms for a frame t using the BoW. We determine the 20 best matching frames $M = \{m_1, ..., m_{20}\} \subset F_{train}$ with the training set by comparing histograms through histogram intersection. Afterwards, we directly compare the SURE features $S(t)$ in frame t with the features $S(m_i)$ in each of the 20 best matching frames $m_i \in M$. Our error metric sums up the minimal distance of pairings of each feature in $S(t)$ with all features in $S(m_i)$.

$$D(S(t), S(m_i)) = \sum_{s \in S(t)} \min_{s_M \in S(m_i)} d(s, s_M) \qquad (4)$$

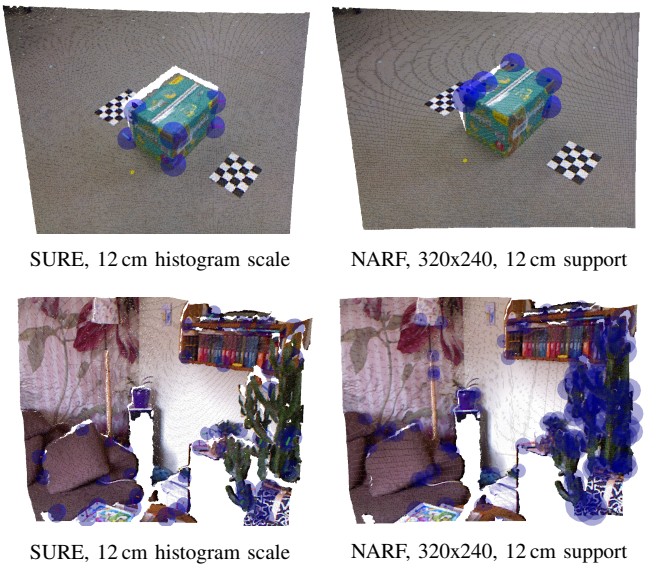
Finally, the semantic label of the best matching frame

$$m_{\text{best}} = \arg\min_{m_i \in M} D(S(t), S(m_i)) \qquad (5)$$

is chosen as the classification of t.

VI. Experiments

We evaluate SURE features and our place recognition approach on RGB-D images from a Microsoft Kinect sensor and compare SURE with the NARF interest point detector and descriptor as implemented in PCL 1.4 (cf. [28]). In all experiments, we evaluate SURE at full resolution (640×480).



SURE, 12 cm histogram scale | NARF, 320x240, 12 cm support

SURE, 12 cm histogram scale | NARF, 320x240, 12 cm support

Fig. 7: Examples of detected interest points on a box (top) and in a living room scene (bottom).

A. Repeatability of the Detector

We assess the quality of our interest point detector by measuring its repeatability across view-point changes. We recorded 4 scenes, 3 containing objects of various size, shape, and color, and one cluttered scene with many objects in front of a wall. The objects are a box (ca. 50x25x25 cm^3), toy rocking horses (height ca. 1 m), and a teddy bear (height ca. 20 cm). Image sequences with 80 to 140 VGA images (640×480 resolution) have been obtained by moving the camera around the objects. We estimate the ground truth pose of the camera using checkerboard patterns laid out in the scenes.

In each image of a sequence, we extract interest points on 3 histogram scales (SURE) or support sizes (NARF). We chose the scales 12, 24, and 48 cm. We then associate interest points between each image pair in the sequence using the ground truth transform. Each interest point can only be associated once to an interest point in the other image. We establish the mutually best correspondences according to the Euclidean distance between the interest points. Valid associations must have a distance below the histogram scale (SURE) or support size (NARF) of the interest point. In addition to this simple measure of repeatability, we also propose a measure that takes into account the uniqueness of the interest point. This "unique repeatability" only accepts an association if it is unambiguous, that means, if the associated interest point is the only one within the acceptance volume. If two or more interest points are within the accepted distance, an association will be rejected.

While the simple repeatability measure (Fig. 8, mid row) seems to indicate that NARF performs better than SURE, our features clearly outperform NARF in unique repeatability (Fig. 8, top row). This is due to the fact that NARF finds many possibly redundant interest points. Thus, SURE provides features that can be uniquely matched between frames,

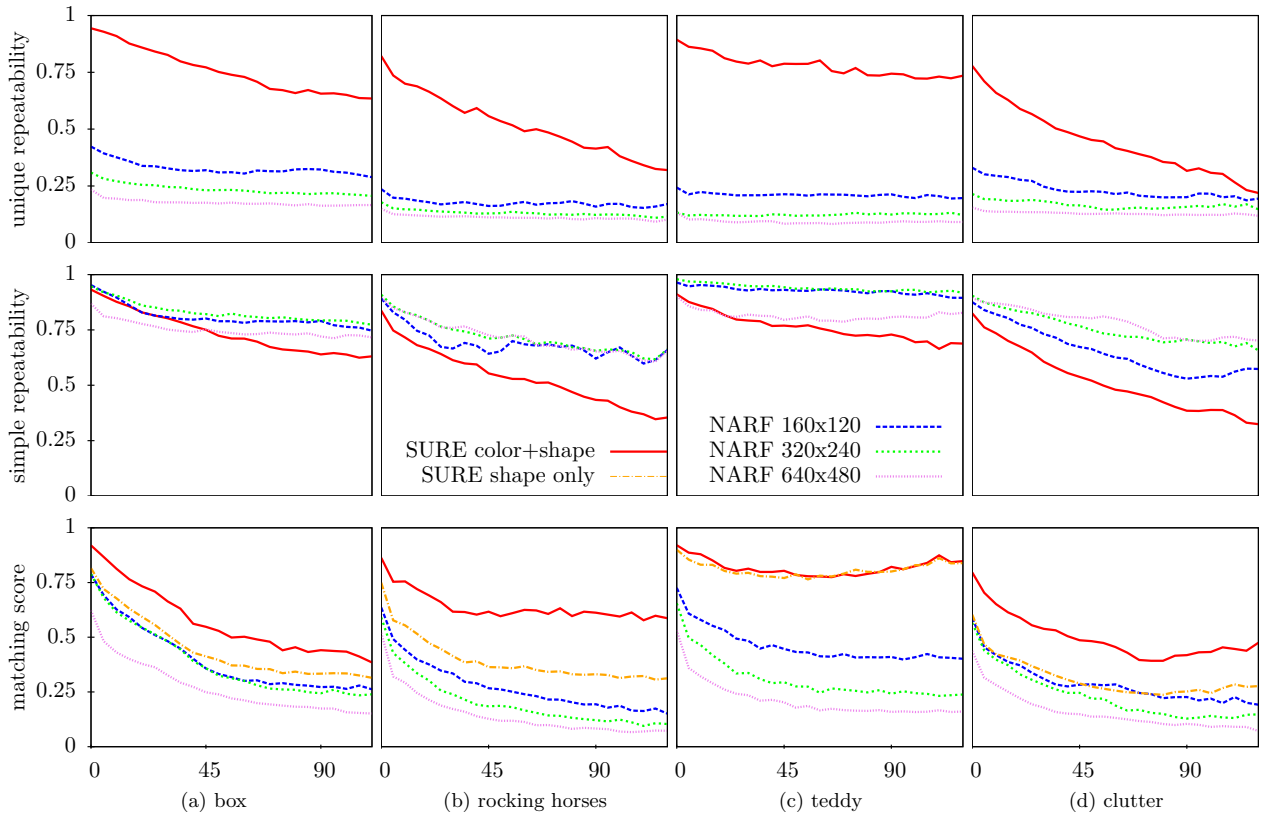


Fig. 8: Repeatability and matching score of SURE and NARF features in different scenes under view-point change (x-axis, in degrees). Top row: Repeatability of unique interest points for which a repetition must be unambiguous in its support volume. Mid row: Simple repeatability where each interest points is mapped to the closest interest point in the second image. Bottom row: matching repeatability of the descriptors referring to the closest interest points in the second image.

which reduces the chance of false associations.

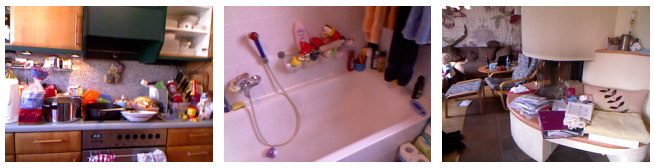
B. Matching Score of the Descriptor

We also evaluate the capability of the detector-descriptor pair for establishing correct matches between images. We define the matching score as the fraction of interest points that can be correctly matched between images by descriptor.

The results in the bottom row of Fig. 8 clearly demonstrate that SURE performs better than NARF in matching individual interest points. In all resolutions NARF detects more interest points than SURE (see Table I). The NARF descriptor which only incorporates shape information does not seem to be distinctive enough to reliably find correct matches. SURE with only shape information performs similar or better in matching score. For the teddy scene, shape seems to be the essential feature, which is well captured by SURE with and without colorful texture in its descriptor. SURE focuses on prominent local structure that is well distinguishable with our descriptor, and it can take advantage of color and luminance information.

C. Run-Time

Table I lists the run-times of SURE and NARF on the 4 datasets. NARF shows significant differences in the average run-time depending on the resolution.



(a) kitchen scene (b) bathroom scene (c) living room scene

Fig. 11: Example images of the second location (3 rooms).

D. Place Recognition Results

For the evaluation of our place recognition approach, we recorded two locations with multiple scans of different rooms with the Kinect. The first location contains five different rooms (a living room, bathroom, kitchen, bedroom and corridor) with approx. 200 frames overall, the second location contains three different rooms with approx. 500 frames (see Fig. 11). For every location two sets were recorded to gain independent data for training and testing.

The results are displayed in Table II. Our approach classifies nearly 90% of frames correctly. The direct comparison of SURE features (i.e., comparing the test set to all images in the training data set according to Eq. (5)) is only slightly better, but needs more computational effort to classify each frame. The results of NARF and SURE without RGB in-

	SURE		NARF 160x120		NARF 320x240		NARF 640x480	
dataset	#features	run-time (sec)	#features	run-time (sec)	#features	run-time (sec)	#features	run-time (sec)
box	8.8	0.62	14.8	0.27	18.2	1.95	32.5	160.18
rocking horses	19.8	0.72	44.6	0.36	72.4	3.25	121.6	133.36
teddy	3.9	0.72	15.3	0.26	26.9	2.09	43.0	164.43
clutter	26.4	0.84	26.5	0.27	48.4	3.24	93.3	179.20

TABLE I: Average number of features and average run-time in seconds per frame.

dataset	method	correct	avg. run-time
3 rooms	NARF 160x120	53,3%	0.3 sec
	NARF 320x240	55,1%	4 sec
	NARF 640x480	28,1%	4 min
	SURE shape only	56,2%	1.4 sec
	SURE color+shape	91,5%	1.6 sec
5 rooms	NARF 160x120	37,5%	0.3 sec
	NARF 320x240	39,1%	4 sec
	NARF 640x480	24,0%	4 min
	SURE shape only	43,4%	1.4 sec
	SURE color+shape	88,3%	1.9 sec
	SURE, direct comparison	91,1%	4.8 sec

TABLE II: Place recognition results.The average run-time includes the time needed for creating the features and the place recognition task averaged per frame.

formation indicate that the use of shape alone for feature description is not sufficient to distinguish places with our approach.

VII. Conclusions

We proposed SURE, a novel pair of interest point detector and descriptor for 3D point clouds and depth images, and applied it for place recognition. Our interest point detector is based on a measure of surface entropy on normals that selects points with strong local surface variation. We designed a view-pose-invariant descriptor that quantifies this local surface curvature using surfel-pair relations. When RGB information is available, we also incorporate colorful texture.

In experiments, we could demonstrate that the SURE descriptor outperforms NARF in matching corresponding features. While NARF finds many redundant interest points in an image, SURE detects more distinctive features. SURE also performs faster than NARF on 640×480 images. We could also demonstrate that SURE features are well suited for place recognition using a Bag-of-Words approach.

In future work, we will apply place recognition based on SURE for loop-closure detection in a 3D SLAM framework. We will also investigate automatic scale selection to further improve the repeatability and localization of the SURE interest points.

References

[1] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.

[2] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. Journal of Computer Vision*, 2004.

[3] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," 2002.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, no. 2, p. 91, 2004.

[5] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," 2006.

[6] T. Kadir and M. Brady, "Saliency, scale and image description," *Int'l J. of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.

[7] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," 2004.

[8] W.-T. Lee and H.-T. Chen, "Histogram-based interest point detectors," 2009.

[9] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[10] P. Moreels and P. Perona, "Evaluation of feature detectors and descriptors based on 3D objects," *Int. Journal of Computer Vision*, 2007.

[11] J. Novatnack and K. Nishino, "Scale-dependent 3D geometric features," in *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, 2007.

[12] R. Unnikrishnan and M. Hebert, "Multi-scale interest regions from unorganized point clouds," in *Workshop on Search in 3D, IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[13] J. Stückler and S. Behnke, "Interest point detection in depth images through scale-space surface analysis," in *Proc. of the IEEE Int. Conference on Robotics and Automation (ICRA)*, 2011.

[14] B. Steder, G. Grisetti, and W. Burgard, "Robust place recognition for 3D range data based on point features," in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2010.

[15] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in *Proc. of the European Conference on Computer Vision (ECCV)*, 2010.

[16] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," 2011.

[17] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 21, no. 5, 1999.

[18] A. Frome, D. Huber, R. Kolluri, T. Blow, and J. Malik, "Recognizing objects in range data using regional point descriptors," 2004.

[19] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 27, no. 11, pp. 1832 –1837, nov. 2005.

[20] F. Tombari, S. Salti, and L. di Stefano, "A combined texture-shape descriptor for enhanced 3D feature matching." in *Proc. of the IEEE Int. Conference on Image Processing (ICIP)*, 2011.

[21] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," 2009.

[22] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *Int. Journal of Robotics Research*, 2010.

[23] B. Steder, M. Ruhnke, S. Grzonka, and W. Burgard, "Place recognition in 3D scans using a combination of bag of words and point feature based relative pose estimation," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2011.

[24] T. R. Shah, "Automatic reconstruction of industrial installations using point clouds and images," Ph.D. dissertation, TU Delft, 2006.

[25] E. Wahl, G. Hillenbrand, and G. Hirzinger, "Surflet-pair-relation histograms: a statistical 3D-shape representation for rapid classification," in *Proc. of the Int. Conf. on 3-D Digital Imaging and Modeling*, 2003.

[26] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. of Computer Vision*, 2000.

[27] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *Proc. of the Int. Conference on Computer Vision (ICCV)*, 2009.

[28] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," Shanghai, China, May 9-13 2011.